MULTIVARIATE WIDE-SENSE MARKOV PROCESSES AND PREDICTION THEORY[1]

By Frederick J. Beutler

University of Michigan

1. Introduction. This paper analyzes multivariate wide-sense Markov processes (hereafter abbreviated MWM processes) as a generalization of the univariate case (see [4], p. 233). Known properties are extended to MWM processes and new results are obtained; these are useful in the theory of multivariate prediction.

Of MWM processes, only the stationary multivariate gaussian Markov process appears to have been studied [3]. Analysis of this process has centered on its second-order properties, the gaussian density function serving as a principal tool [7]. Actually, the same second-order properties hold for any MWM, and these depend only on the covariance matrix. The special character of this matrix (Theorems 2 and 3) is therefore the key to the analysis of MWM processes.

The most significant application of MWM processes is to the theory of linear minimum-mean-square-error prediction of (not necessarily wide-sense stationary) multivariate stochastic processes. Ordinarily, the optimum prediction may be determined by solving simultaneous sets of Fredholm integral equations of the first kind—a rather hopeless task.[2] For MWM processes, however, prediction problems admit of explicit solutions determined almost by inspection. The solution technique is readily extended to linear functionals on MWM processes, using Bochner integrals to supply the analytic basis.

Without attempting to be precise at this time, we can nevertheless illustrate the scope of generalized prediction of MWM. Consider the integral equation

$$x(t) = \int_0^t A(\tau)x(\tau)\,d\tau + \int_0^t M(\tau)\,dy(\tau), \qquad t \geqq 0, \tag{1.1}$$

in which upper- and lower-case symbols represent matrices and vectors, respectively. If $y(\cdot)$ is a second-order temporally and spatially homogeneous (vector) process with orthogonal increments, and if some integrability conditions are satisfied, (1.1) may be regarded as an integral equation defined on an abstract function space and possessed of a unique solution. Indeed, the solution $x(\cdot)$ is exhibited and shown to constitute a MWM process. Consequently, the linear mean-square optimum predictor for $x(t + \alpha)$, $\alpha > 0$, based on $x(s)$, $t - T \leqq s \leqq t$, can be explicitly formulated. We can also find linear optimum

Received May 4, 1962; revised January 30, 1963.

[1] This work was supported by the National Aeronautics and Space Administration under research grant NsG-2-59.

[2] Indeed, solutions may not exist, or may appear as distributions rather than functions.

estimates of $z(t) = \int_{-\infty}^{\infty} K(t, s)x(s)\, ds$, where $x(\cdot)$ is as in (1.1), and the estimate is based on the availability of $x(\cdot)$ over a specified subset of the reals.

2. Definitions. A multivariate second-order process $x(\cdot)$ consists of n component processes $x_1(\cdot), x_2(\cdot), \cdots, x_n(\cdot)$, each of which is itself a complex second-order process. It will be convenient throughout to regard $x(t)$ as the matrix

$$\begin{bmatrix} x_1(t) & 0 & 0\cdots 0 \\ x_2(t) & 0 & 0\cdots 0 \\ & \cdots & \\ x_n(t) & 0 & 0\cdots 0 \end{bmatrix} \tag{2.1}$$

and $x^*(t)$ as the complex conjugate transpose of $x(t)$.

The expectation of a matrix is the matrix formed of the expectation of each element of the original matrix. For instance, $E[x(s)x^*(t)]$ is the matrix whose ij element is $E[x_i(s)x_j^*(t)]$. An analogous convention applies to other linear scalar operations such as integration, differentiation, and scalar multiplication

We shall now construct a Hilbert space H as follows. If the prefix τ denotes the trace of a matrix,

$$(x(s), x(t)) = \tau\{E[x(s)x^*(t)]\} \tag{2.2}$$

is an inner product which is a positive symmetric bilinear functional. However, this inner-product space is insufficiently large for our purposes; we need to consider all elements of the form $A(t)x(t)$, where $A(\cdot)$ may be any $n \times n$ matrix with (finite) complex nonstochastic entries. The inner product of two such elements, say $A(s)x(s)$ and $B(t)x(t)$, is then

$$(A(s)x(s), B(t)x(t)) = \tau\{A(s)(E[x(s)x^*(t)])B^*(t)\}. \tag{2.3}$$

The space of elements such as $A(t)x(t)$ is extended to a linear manifold which is completed to H, the Hilbert space we shall work in henceforth.

Next, we define the wide-sense expectation $\hat{E}$; for any $z \varepsilon H$, and any subset of the real line, $\hat{E}\,[z \mid x(t), t \varepsilon T]$ is the projection of z on the subspace M, where M is the span

$$M = V\{A(t)x(t) : t \varepsilon T, \text{all finite valued } A(t)\}. \tag{2.4}$$

In terms of prediction theory, $\hat{E}[z \mid x(t), t \varepsilon T]$ is the optimum linear estimate (in M) of z, the optimization criterion being the sum of the mean-square component errors. In other words, we have minimized

$$\mathcal{E}^2 = \|z - y\|^2 = \sum_1^n [E \mid z_j - y_j \mid^2] \tag{2.5}$$

for a specified $z \varepsilon H$ over all $y \varepsilon M$. That such a $y \varepsilon M$ exists and is unique (up to a set of probability zero) is a well-known result applicable to projection operators in a Hilbert space. The minimum property of y is equivalent to the orthogonality condition $(z - y) \perp A(t)x(t)$ for all finite valued $A(t)$ and $t \varepsilon T$. Orthogonality is usually most easily verified by the criterion of[3]

LEMMA 1. *Given a random variable w and any invertible matrix C, $w \perp M$ if and only if $E[wx^*(t)C] = 0$ for all $t \varepsilon T$.*

The proof of Lemma 1 is obvious. Usually, we take C to be I (the identity matrix) in the application of this criterion.

The error criterion (2.5) may prove inappropriate for some purposes. We shall therefore consider

$$(2.6)\quad \mathcal{E}^2(t) = \sum_i \{E \; |\sum_k b_{ik}(t)[z_k(t) - \tilde{y}_k(t)]|^2\} = \|B(t)[z(t) - \tilde{y}(t)]\|^2$$

as a generalized figure of merit, with $B(t)$ any finite complex matrix, and $\tilde{y}(t)$ an estimate of $z(t)$. We assert that the optimization solution already discussed specifies the $y(t) \varepsilon M$ which minimizes (2.6). Indeed, we have the stronger result of

THEORM 1. *If $y(t) \varepsilon M$ minimizes* (2.5), *$\tilde{y}(t) = y(t)$ minimizes* (2.6) *over all $\tilde{y}(t) \varepsilon M$. If, in addition, $B(t)$ is nonsingular (at t), $\tilde{y}(t) = y(t)$ is the unique element of M minimizing* (2.6), *up to sets of probability zero.*

PROOF. If we let $w(t) = B(t)\tilde{y}(t)$, the problem of minimizing (2.6) is reduced to finding the unique $w(t) \varepsilon M_1$ which minimizes $\|B(t)z(t) - w(t)\|$. Here M_1 is

$$(2.7)\qquad M_1 = V\{A(t)B(t)x(t) : t \varepsilon T, \text{ all finite valued } A(t)\}$$

which is the subspace containing $w(t)$ corresponding to $y(t) \varepsilon M$. Evidently, $M_1 \subset M$; if $B(t)$ is invertible (i.e., nonsingular), $M_1 = M$.

We now show that the choice $w(t) = B(t)y(t)$ yields $[B(t)z(t) - w(t)] \perp M_1$. Indeed, if $u(t) \varepsilon M_1 \subset M$, $B^*(t)u(t) \varepsilon M$ [see (2.4)], and so

$$(2.8)\qquad ([B(t)z(t) - w(t)], u(t)) = ([z(t) - y(t)], B^*(t)u(t)) = 0$$

because $y(t) \varepsilon M$ minimizes (2.5). Since the $w(t)$ satisfying (2.8) [and thus minimizing (2.6)] is unique up to a set of zero probability,

$$(2.9)\qquad B(t)\tilde{y}(t) = B(t)y(t) \text{ probability 1 for each } t$$

must be satisfied if (2.6) is to be minimized. It is clear that $\tilde{y}(t) = y(t)$ satisfies this equation in any case. If $B(t)$ is invertible, $\tilde{y}(t) = y(t)$ with probability one for each t is the only solution of (2.9). Thus the proof of the theorem is complete.

We wish to define a normalized covariance $R(\cdot\,,\cdot)$ which possesses the property

$$(2.10)\qquad \hat{E}[x(t) \mid x(s)] = R(t, s)x(s).$$

If we call $P(s, t)$ the covariance matrix whose ij entry is

$$(2.11)\qquad P_{ij}(s, t) = E[x_i(s)x_j^*(t)]$$

[3] This concept of orthogonality is stronger than the customary motion. For $y \perp z$, it is not sufficient that $(y, z) = 0$.

it is easily verified from Lemma 1 (with $C = I$) that

$$R(t, s) = P(t, s)[P(s, s)]^{-1} \tag{2.12}$$

satisfies the required orthogonality condition provided $P(s, s)$ is invertible. We are more interested, however, in the generalization of (2.12) which does not require $P(s, s)$ to be nonsingular. Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be the eigenvalues of $P(s, s)$ and take D as the diagonal matrix whose entries are $d_{ij} = \lambda_i\delta_{ij}$ (note that s is suppressed in this discussion). Since $P(s, s)$ is Hermitian, n orthogonal eigenvectors $\{c_{i1}, c_{i2}, \cdots, c_{in}\}$ correspond to the eigenvalues; we may take these to be normalized by $\sum_j | c_{ij} |^2 = 1$. The unitary matrix C is then formed from the elements c_{ij}.

Finally, let E be the diagonal matrix with $e_{ii} = \lambda_i$ if $\lambda_i \neq 0$ and $e_{ii} = 1$ if $\lambda_i = 0$. Then

$$R(t, s) = P(t, s)[C^*E^{-1}C] \tag{2.13}$$

satisfies (2.10), as we shall show. If D is nonsingular, (2.12) and (2.13) are the same, and the result is immediate. More generally, we must show that $R(t, s)$ given by (2.13) satisfies $[x(t) - R(t, s)x(s)] \perp Ax(s)$ for arbitrary A. According to Lemma 1, the orthogonality requirement is the equivalent of

$$P(t, s)C^* = R(t, s)P(s, s)C^* \tag{2.14}$$

which we proceed to verify. First, $P(s, s)C^* = C^*D$, so that $E^{-1}CP(s, s)C^* = E^{-1}D$. Therefore, the right side of (2.14) becomes $P(t, s)C^*(E^{-1}D)$. Here $E^{-1}D$ is the diagonal matrix whose entries are 1 or 0 according as the corresponding $\lambda_i \neq 0$ or $\lambda_i = 0$. A comparison of the right and left sides of (2.14) then shows equality to hold if the ith column of $P(t, s)C^*$ is zero whenever $\lambda_i = 0$. The latter is proved as follows. Since $CP(s, s)C^*$ and $E\{[Cx(s)][Cx(s)]^*\}$ are the same, we find that

$$E[|\sum_j c_{ij}x_j(s)|^2] = \lambda_i \tag{2.15}$$

whence $\sum_j c_{ij}x_j(s) = 0$ (probability 1) for all i such that $\lambda_i = 0$. This means that $x^*(s)C^*$ has zeros in its ith column with probability 1 whenever $\lambda_i = 0$. Consequently, so does $P(t, s)C^* = E[x(t)x^*(s)C^*]$, and the proof is complete.

We conclude this section with the remark that the conditional wide-sense expectation of any random variable is defined only with probability one. Such a projection ought therefore to be reviewed as any one of an equivalence class of random variables differing from one another only on sets of zero probability. Since these sets of probability zero do not, in general, play an essential role in our theory, the phrase "with probability 1" will usually be omitted.

3. Wide-sense Markov processes. A process is defined to be MWM (multivariate wide-sense Markov) if

$$\hat{E}[x(u) \mid x(s), x(t)] = \hat{E}[x(u) \mid x(t)] \tag{3.1}$$

whenever $s \leqq t \leqq u$. With the assumption

$$\lim_{s-t \to 0} \|x(s) - x(t)\| = 0, \tag{3.2}$$

the MWM property has the alternative characterization

$$\hat{E}[x(t) \mid x(s), s \varepsilon T] = \hat{E}[x(t) \mid x(s^*)], \quad s^* = \sup_{s \varepsilon T} s \leqq t. \tag{3.3}$$

Note that $x(s^*) \varepsilon M$ by (3.2), and that (3.1) assures that $[x(t) - R(t,s^*)\, x(s^*)] \perp A(s)x(s)$, any finite $A(s)$, $s \varepsilon T$.

The prediction problem is easily solved for MWM processes. If the mean-square linear prediction to $x(t + \alpha)$, $\alpha > 0$, is to be based on $x(s)$, $s \leqq t$, the desired estimate is seen to be $R(t + \alpha, t)x(t)$.

In general, the wide-sense Markov property is difficult to verify directly. However, there is an explicit description of such processes in terms of $R(s, t)$. The principal result in this direction is

THEORM 2. *The following statements are all equivalent:*

(1) $x(\cdot)$ *is wide-sense Markov.*

(2) *Whenever* $s \leqq t \leqq u$,

$$R(u, s) = R(u, t)R(t, s). \tag{3.4}$$

(3) *Whenever* $s \leqq t \leqq u$,

$$R(s, u) = R(s, t)R(t, u). \tag{3.5}$$

(4) $x(- \cdot)$ *is wide-sense Markov.*

PROOF. We will show that each statement implies the succeeding one. If (1) holds with $s \leqq t \leqq u$, $\hat{E}[x(u) \mid x(t), x(s)] = R(u, t)x(t)$ from (3.1) and (2.10). From the orthogonality relation and Lemma 1, $E\{[x(u) - R(u, t)x(t)]x^*(s)\} = 0$; this may also be written as $P(u, s) = R(u, t)P(t, s)$. Finally, both sides of the preceding equation are post-multiplied by $C^*E^{-1}C$ to give the desired result (3.4).

Next, let us assume that (2) is true. We adopt the notational convenience $F(t) = C^*(t)E(t)^{-1}C(t)$ and observe that the matrix so defined is Hermitian. Then $R^*(s, t) = F(t)P(t, s)$ follows from $P^*(s, t) = P(t, s)$. Substituting for R^* in the equation $R^*(u, s) = R^*(t, s)R^*(u, t)$ enables us to obtain the result

$$F(s)P(s, u) = F(s)P(s, t)F(t)P(t, u). \tag{3.6}$$

Since $F(s)$ is invertible, (3.6) may be pre-multiplied by $F(s)^{-1}$ and postmultiplied by $F(u)$ to complete the proof of (3).

As a consequence of (3),

$$R(-u, -s) = R(-u, -t)R(-t, -s) \tag{3.7}$$

since $-u \leqq -t \leqq -s$. Now if $\tilde{x}(t) = x(-t)$, $\tilde{x}(\cdot)$ has a normalized covariance $\tilde{R}$ given by $\tilde{R}(s, t) = R(-s, -t)$. Then (3.7) becomes

$$\tilde{R}(u, s) = \tilde{R}(u, t)\tilde{R}(t, s); \tag{3.8}$$

this, as we shall demonstrate, makes $\tilde{x}(\cdot)$ a MWM process. In fact, (3.8) implies that

$$E\{[\tilde{x}(u) - \tilde{R}(u,t)\tilde{x}(t)]\tilde{x}^*(s)\tilde{F}(s)\} = 0, \tag{3.9}$$

for $s \leqq t \leqq u$. By Lemma 1, then, $\tilde{x}(u) - \tilde{R}(u,t)\tilde{x}(t)$ is orthogonal to a subspace which includes both $\tilde{x}(s)$ and $\tilde{x}(t)$ [set $s = t$ in (3.9)]. Hence

$$\hat{E}[\tilde{x}(u) \mid \tilde{x}(s), \tilde{x}(t)] = \tilde{R}(u,t)\tilde{x}(t) \tag{3.10}$$

the right side of which is equal to $\hat{E}[\tilde{x}(u) \mid \tilde{x}(t)]$ from (2.10). This means that $\tilde{x}(t) = x(-t)$ satisfies the definition of a MWM process.

Thus this succession of steps shows that (1) implies (4). Because of symmetry considerations, an identical procedure verifies that (4) implies (1), completing the proof of the theorem.

If further assumptions are made about the nature of $x(\cdot)$, even more can be said of the normalized covariance of the process. It is known, for instance, that if $x(\cdot)$ is a multivariate nonsingular stationary gaussian Markov process, $R(s,t)$ takes on the form $R(s,t) = e^{C(s-t)}$ for $s > t$, C being some constant matrix. We shall prove a similar theorem, obtaining stronger results from weaker assumptions.

We shall suppose that $x(\cdot)$ is wide-sense stationary, which means that $E[x_i(s)x_j^*(t)]$ is a function of $s - t$ only for every s, t, i, and j. In other words, we may write (in somewhat questionable but self-explanatory notation) $P(s,t) = P(s-t)$. If we suppose further that $P(0)$ is invertible,

$$R(s,t) = R(s-t) = P(s-t)[P(0)]^{-1}. \tag{3.11}$$

We shall also give $x(\cdot)$ certain continuity properties, which are completely analogous to continuity in norm for a univariate process. Indeed, if the multivariate $x(\cdot)$ is continuous in norm from the right (or left) at any t, $x(\cdot)$ is continuous in norm everywhere. This assumption also suffices to render $P(\cdot)$ and $R(\cdot)$ continuous. Continuity of these matrices may be either in norm ($\|A\|^2 = \tau(AA^*)$) or individually for every matrix element; for any matrix, these two notions of continuity are equivalent. Conversely, if $P(\cdot)$ or $R(\cdot)$ is continuous from the right (left) at the origin, $P(\cdot)$ and $R(\cdot)$ are continuous everywhere, and $x(\cdot)$ is continuous in norm.

Under the assumptions we have just made, the MWM property is completely characterized by

THEOREM 3.[4] *Let $x(\cdot)$ be wide-sense stationary and continuous in norm with a nonsingular $P(0)$. Then $x(\cdot)$ is* MWM *if and only if*

$$R(t) = e^{Ct} \qquad t \geqq 0 \tag{3.12}$$

where C is a constant matrix none of whose eigenvalues has positive real parts.

[4] Only the second statement of the theorem is new. The first has been given various proofs [3], [7]. Doob's proof [3] is based on matrix theory, but differs from ours.

REMARK. For $t < 0$, the wide-sense stationary MWM process has $R(t) = e^{-C^*t}$.

PROOF. Since an $R(\cdot)$ of the form (3.12) satisfies (3.4), the "if" part of the theorem is proved. To prove "only if," observe first that (3.4) and the stationarity property give

$$R(t + s) = R(t)R(s) \qquad t, s \geqq 0. \tag{3.13}$$

By finite induction, this equation leads to $R(mt) = [R(t)]^m$. Hence $[R(s/n)]^n = R(s)$ and, since roots of matrices are well defined, $R(s/n) = [R(s)]^{1/n}$. Upon setting $t = 1/n$ and $s = 1$, we find that

$$R(r) = [R(1)]^r \qquad r \text{ rational}, r \geqq 0. \tag{3.14}$$

The continuity of R extends (3.14) to nonrational t, so that $R(t) = [R(1)]^t$ for all $t \geqq 0$.

We would like to assert that there exists a matrix C such that $R(1) = e^C$. There is such a matrix if and only if $R(1)$ is nonsingular (cf. [1], p. 29). But in fact, $R(t)$ is nonsingular for every $t \geqq 0$ (3.13). In the first place, $R(0) = I$ is nonsingular. Then, by the continuity of R there is a $\delta > 0$ such that $R(t)$ is nonsingular for $0 \leqq t \leqq \delta$. Now let $t_0 > \delta$ be the smallest value of t for which $R(t)$ is singular. Then $R(t_0) = R(t_0 - \delta)R(\delta)$ implies that $R(t_0 - \delta)$ is singular, which contradicts the existence of the assumed t_0.

From the above argument, $R(1) = e^C$, so that $R(t) = [e^C]^t = e^{Ct}$. Only the statement about the eigenvalues of C needs to be proved. This assertion of the theorem is true if none of the eigenvalues of e^C has modulus greater than unity. Suppose that the theorem is false, so that $R(1)$ has an eigenvalue λ with $|\lambda| > 1$. Now $R(n) = [R(1)]^n$; therefore, $R(n)$ has an eigenvalue λ^n, and

$$\limsup_{t\to\infty} [\max_j |\lambda_j(t)|] = \infty. \tag{3.15}$$

But, as we shall show, the eigenvalues of $R(t)$ are uniformly bounded, which completes the proof *ab contrario.*

Using the Schwartz inequality (applied in this case to finite sums), we obtain $||AB|| \leqq ||A|| \cdot ||B||$. We can show similarly that $||P(t)|| \leqq ||x(0)||^2$ for all t. From these considerations it follows that

$$|\lambda(t)| \leqq ||R(t)|| \leqq ||x(0)||^2 \, ||P(0)^{-1}||. \tag{3.16}$$

Thus the contradiction is reached, and the proof is complete.

4. Functionals of MWM processes. We now consider the minimum-mean-square-error linear estimation of the random variable

$$z = \int B(t)x(t) \, dt; \tag{4.1}$$

that is, we wish to obtain $\hat{E}[z \mid x(t), t \in T]$. It is hoped that this wide-sense expectation is given by

$$\hat{E}[z \mid x(t), t \varepsilon T] = \int B(t)\hat{E}[x(t) \mid x(s), s \varepsilon T] \, dt. \tag{4.2}$$

If (4.2) is true, and $x(\cdot)$ is MWM, it should be easy to calculate $\hat{E}[z \mid x(t), t \epsilon T]$.

In view of (3.2), there is a version of $x(\cdot)$ measurable with respect to the product measure constituted of Lebesgue measure on the real line and probability measure on an abstract space. Thus the integral can be defined for almost all sample functions according to the procedure of [4], pp. 61–63. A separate argument would still be needed to prove $z \epsilon H$ and (4.2). To avoid some of these arguments, and to provide a more elegant structure, we shall define (4.1) as a Bochner integral instead. This approach yields the same stochastic integral (up to an equivalence) as [4] when the latter is defined.

It is hereafter assumed that each element of $B(\cdot)$ is (Lebesgue) measurable, that $x(\cdot)$ satisfies (3.2), and that

$$\int_{-\infty}^{\infty} \|B(t)x(t)\| \, dt < \infty. \tag{4.3}$$

Under these conditions, the Bochner integral (4.1) is shown to exist, z belongs to H, and (4.2) is valid.

In the first place, the (Lebesgue) measurability of each element of $B(\cdot)$ implies that $B(\cdot)$ is strongly measurable according to [6], Definition 3.5.4. Since $x(\cdot)$ is continuous in norm, it too is strongly measurable. Then the product $B(\cdot)x(\cdot)$ is likewise strongly measurable. Consequently, $\|B(\cdot)x(\cdot)\|$ is (Lebesgue) measurable and, under assumption (4.3), $B(\cdot)x(\cdot)$ is Bochner integrable (cf. [6], Theorem 3.7.4). It is evident from the construction of this integral that $z \epsilon H$. Finally, $\hat{E}$ is a bounded linear transformation of H into itself, so that Theorem 3.7.12 of [6] may be employed to yield (4.2).

We show that, whenever (4.1) exists as the integral described in [4], pp. 51–63, this integral coincides with the Bochner integral up to the usual equivalence. To avoid any confusion in the proof, we shall adopt the notations (B) $\int$ and (L) $\int$ to designate the Bochner or Lebesgue integral, respectively.

The proof is accomplished if there is for each n a countably valued function $B^{(n)}(t)x^{(n)}(t)$ for which both integrals exist, with

$$(\mathrm{B}) \int B^{(n)}(t)x^{(n)}(t) \, dt = (\mathrm{L}) \int B^{(n)}(t)x^{(n)}(t) \, dt \quad (\text{prob. } 1) \tag{4.4}$$

and the additional properties

$$\left\| (\mathrm{B}) \int B^{(n)}(t)x^{(n)}(t) \, dt - (\mathrm{B}) \int B(t)x(t) \, dt \right\| < 1/n \tag{4.5}$$

and

$$E\left| (\mathrm{L}) \int b_{ij}^{(n)}(t)x_j^{(n)}(t) \, dt - (\mathrm{L}) \int b_{ij}(t)x_j(t) \, dt \right| < 1/n. \tag{4.6}$$

There exist measurable disjoint sets $\{E_k\}$, fixed $t_k \in E_k$, and countably valued functions

$$B^{(n)}(t)x^{(n)}(t) = \sum_k I_k(t)B(t_k)x(t_k) \tag{4.7}$$

such that (cf. [6], Corollary to Theorem 3.7.4)

$$\int \|B^{(n)}(t)x^{(n)}(t) - B(t)x(t)\| \, dt < 1/n. \tag{4.8}$$

Here $I_k(t)$ is the identity matrix or zero according as $t \in E_k$ or not. For the function described by (4.7),

$$(\mathrm{B}) \int B^{(n)}(t)x^{(n)}(t) \, dt = \sum_k B(t_k)x(t_k)m(E_k), \tag{4.9}$$

$m(E_k)$ being the Lebesgue measure of E_k. The right side of (4.9) is properly defined as a random variable. Indeed, this expression is a sum of random variables; furthermore, the sum is absolutely convergent with probability one from the inequality

$$E \sum_k |b_{ij}(t_k)x_j(t_k)| \, m(E_k) < \int \|B(t)x(t)\| \, dt + 1/n. \tag{4.10}$$

The function (4.7) is also (L)-integrable for almost all sample functions. Since (4.7) is a countable sum of products of Lebesgue- and probability-measurable functions, it is itself jointly measurable. To complete the argument, we use (4.10) to show that the conditions of [4], Theorem 2.7, p. 62, are satisfied. The resulting integral will then be identical with the right side of (4.9), except possibly on a set of zero probability.

It is now easy to verify (4.5) and (4.6). Applying the norm inequality of [6], Theorem 3.7.6, to (4.8) shows (4.5) to be true. (4.6) is successively majorized by taking absolute values of the integrand, and applying Fubini's theorem and Jensen's inequality. The result is, in turn, majorized by (4.8), so that the proof is complete.

The results obtained thus far may be summarized in

THEOREM 4. *Let $B(\cdot)$ be measurable, and let* (3.2) *and* (4.3) *be satisfied. Then* (4.1) *may be defined as a Bochner integral, and z may be regarded as the* (*unique up to an equivalence*) *random variable obtained as the limit in H norm of* (4.9). *If* (4.1) *is also defined as in* [4], *pp.* 61–63, *the resulting random variable agrees with that defined by the Bochner integral, up to the usual equivalence. Furthermore, $z \in H$, and the operator $\hat{E}$ commutes with Bochner integration, as indicated by* (4.2).

In some applications we discuss $z(\cdot) = \int K(\cdot, \tau)x(\tau)\, d\tau$ in place of the random variable z. While conditions on $K(\cdot, \cdot)$ could be given to make $z(\cdot)$ a random process, we are interested only in indexing random variables by the linear parameter set t. Therefore, our only requirement on $K(\cdot, \cdot)$ is that $\int \|K(t, \tau)x(\tau)\| \, d\tau < \infty$ for each t, with $K(t, \cdot)$ measurable in τ for each t.

While the stochastic integral (4.1) is of intrinsic interest, its principal application here will be to the computation of linear minimum-mean-square-error estimates of $z(t) = \int K(t, \tau)x(\tau)\, d\tau$ for MWM $x(\cdot)$. As we have observed before, a principal advantage of dealing with MWM is the ease with which such estimates can be calculated.

Let us suppose that the estimate is to be based on a knowledge of $x(\cdot)$ over the interval $[a, b]$, i.e., $T = [a, b]$. Then we obtain from (4.2), (3.3), and (4) of Theorem 2

$$\hat{E}[z(t) \mid x(s), s \varepsilon T] = \left\{\int_{-\infty}^{a} K(t, \tau)R(\tau, a)\, d\tau\right\} x(a) + \int_{a}^{b} K(t, \tau)x(\tau)\, d\tau + \left\{\int_{b}^{\infty} K(t, \tau)R(\tau, b)\, d\tau\right\} x(b). \tag{4.11}$$

In particular, the last term in (4.11) follows from the fact that

$$\hat{E}[x(t) \mid x(s), s \epsilon T] = \hat{E}[x(t) \mid x(s_*)] = R(t, s_*)x(s_*) \text{ whenever } t \leqq s_* = \inf\{s : s \epsilon T\},$$

which is, in turn, a consequence of (4) of Theorem 2.

The formula (4.11) may be extended and exploited in various ways, some of which lead to engineering applications discussed elsewhere [2].

5. Application to differential equations. In this section we consider the prediction of an $x(t)$ specified by a matrix-vector differential equation with "white noise" input, that is, an equation of the form

$$\dot{x} = A(t)x + M(t)n(t) \qquad x(0) = 0, \qquad t \geqq 0. \tag{5.1}$$

Here $n(\cdot)$ is a multivariate random process whose covariance matrix has elements $q_{ij}(s, t) = \delta_{ij}\,\delta(s - t)$[$\delta_{..}$ and $\delta(\cdot)$ are the Kronecker and Dirac deltas, respectively].

Stochastic differential equations similar to (5.1) have been studied extensively by Ito and others. It is customary to treat the equivalent integral equation

$$x(t) = \int_0^t A(\tau)x(\tau)\, d\tau + \int_0^t M(\tau)\, dy(\tau), \qquad t \geqq 0, \tag{5.2}$$

and to analyze Markov and sample function properties. However, (5.2) may also be viewed—under conditions less restrictive than Ito's—as an equation on an abstract function space. The latter approach appears particularly advantageous when norm topology or linear operator concepts—rather than sample function properties—are of prime concern.

Our assumptions on (5.2) are the following. $y(\cdot)$ is a multivariate process each of whose components is an orthogonal increment process, with

$$E\{[y(t) - y(s)]\,[y(t) - y(s)]^*\} = |t - s|I, \qquad y(0) = 0. \tag{5.3}$$

Then if $\int_0^t \|M(\tau)\|^2\, d\tau < \infty$ for each $t < \infty$, the second integral on the right of (5.2) may be regarded as a stochastic integral, as described in [4], Section 9.2; it is necessary only to extend the usual form to the n-variate case. Finally, it is assumed that the elements of $A(\cdot)$ are measurable and $||A(\cdot)||$ locally integrable, so that the first integral of (5.2) can eventually be defined as a Bochner integral (see Section 4 of this paper).

Under the above hypotheses, there exists (for each $t \geqq 0$) a solution to (5.2) unique up to the usual equivalence. We first prove this only for some neighborhood of the origin $[0, t_1]$, where $t_1 > 0$ is chosen such that $\int_0^{t_1}||A(\tau)||\, d\tau < 1$. Consider a sequence of successive approximations starting with $x_0(\cdot) = 0$ for convenience. Then $x_1(\cdot)$ is continuous in norm; hence, $x_1(\cdot)$ is strongly measurable and bounded (in norm), and $A(\cdot)x_1(\cdot)$ is Bochner integrable on $[0, t_1]$. A process of finite induction shows this to be true also for any $x_n(\cdot)$. Moreover, $\sup_{0 \leqq \tau \leqq t_1} ||x_{n+1}(\tau) - x_{m+1}(\tau)|| \leqq [\sup_{0 \leqq \tau \leqq t_1} ||x_n(\tau) - x_m(\tau)|| \int_0^{t_1}||A(\tau)||\, d\tau$, so that we have a contraction mapping on a Banach space generated by $y(\cdot)$. Hence there is an element $x(\cdot)$ such that $\lim_{n\to\infty} [\sup_{0 \leqq \tau \leqq t} ||x_n(\tau) - x(\tau)||] = 0$. This element, being the uniform limit (in norm) of a sequence of (B)-integrable functions over a finite interval, is itself (B)-integrable, and so satisfies (5.2).

We now extend the solution to any $t < \infty$. For any such t, there exists a finite sequence $\{t_k\}$, $k = 0, 1, \cdots, n$, such that $0 = t_0 < t_1 < \cdots < t_{n-1} \leqq t < t_n$ and $\int_{t_k}^{t_k+1}||A(\tau)||\, d\tau < 1$. Assume that $x(t_k)$ satisfies (5.2) except possibly on a set of zero probability. Then $x(\cdot)$ satisfies (5.2) for $t > t_k$ if and only if it satisfies

$$(5.4) \qquad x(t) = x(t_k) + \int_{t_k}^t A(\tau)x(\tau)\, d\tau + \int_{t_k}^t M(\tau)\, dy(\tau), \qquad t > t_k,$$

in both contexts with probability one for each t. The procedure of the preceding paragraph suffices to assure the existence and uniqueness of solutions for each $[t_{k-1}, t_k]$. Therefore, $x(t)$ is given by the finite sum $x(t) = [x(t) - x(t_{n-1})] + \sum_1^{n-1} [x(t_k) - x(t_{k-1})]$. Since each $x(t_k)$ is defined and unique only up to an equivalence, various versions of $x(t)$ may depend on the choice of the particular t_k as well as that of the $x(t_k)$; nevertheless, two versions of $x(t)$ differ onlv on finite unions of null probability sets.

More specifically, we assert that the solution is provided by the stochastic integral

$$(5.5) \qquad x(t) = \int_0^t H(t, \tau)M(\tau)\, dy(\tau)$$

where where we define

$$(5.6) \qquad H(t, s) = X(t)X^{-1}(s)$$

and $X(\cdot)$ solves the homogeneous matrix differential equation

$$(5.7) \qquad \dot{X} = A(t)X, \qquad X(0) = I.$$

If (5.5) is true, $x(t) \varepsilon H$, where H is generated by $y(\cdot)$. The explicit form (5.5) will also enable us to compute the covariance matrix of $x(\cdot)$, and thus to prove that $x(\cdot)$ is MWM.

We next show that the integral (5.5) exists and satisfies the integral equation

(5.2). To prove existence, we must verify (according to [4], Section 9.2, and our (5.3)) that $\|H(t,\cdot)M(\cdot)\|^2$ is Lebesgue integrable over every $[0, t]$, $t < \infty$. In view of our assumptions about $M(\cdot)$, integrability follows if only $\|H(t,\cdot)\|$ is bounded and measurable. In fact, $X(\cdot)$ and its inverse are continuous (even differentiable). Furthermore, Gronwall's lemma (cf. [1], p. 35) yields

$$\|X(t)\| \leqq n^{\frac{1}{2}} \exp\left\{\int_0^t \|A(\tau)\|\, d\tau\right\} \tag{5.8}$$

where n is the number of components of $x(\cdot)$. The same inequality holds for $X^{-1}(\cdot)$. Thus the right side of (5.5) exists. Indeed, the $x(\cdot)$ so defined is continuous in norm and bounded on each interval $[0, t]$; hence the Bochner integral of $A(\cdot)x(\cdot)$ is well-defined. We may therefore substitute (5.5) into (5.2) to ascertain its validity as a possible solution. When (5.5) is substituted into the first term on the right of (5.2), we obtain from (5.6)

$$\int_0^t A(\tau)x(\tau)\, d\tau = \int_0^t \dot{X}(\tau)\left(\int_0^\tau X^{-1}(\alpha)M(\alpha)\, dy(\alpha)\right) d\tau. \tag{5.9}$$

If integration by parts is valid on the right side of (5.9), (5.2) becomes an identity, so that (5.5) is its solution as claimed.

For notational convenience, let $V(\tau) = \int_0^\tau X^{-1}(\alpha)M(\alpha)\, dy(\alpha)$. Because $V(\cdot)$ is continuous in norm, we may partition $[0, t]$ so that

$$\lim_{\delta\to 0}\left\| \int_0^t X(\tau)V(\tau)\, d\tau - \sum_{i=1}^n [X(\tau_i) - X(\tau_{i-1})]V(u_i) \right\| = 0 \tag{5.10}$$

where $0 = \tau_0 < \tau_1 < \cdots < \tau_n = t$, $\tau_{i-1} \leqq u_i \leqq \tau_i$, and $\delta = \max|\tau_i - \tau_{i-1}|$. But

$$\begin{aligned}\sum_{i=1}^n [X(\tau_i) - X(\tau_{i-1})]V(u_i) &= X(t)V(t)\\ &\quad - \sum_{i=0}^{n-1} X(\tau_i)[V(u_{i+1}) - V(u_i)].\end{aligned} \tag{5.11}$$

The right-hand sum in (5.11) can also be written

$$\sum_{i=0}^{n-1} \int_{u_i}^{u_{i+1}} X(\tau_i)X^{-1}(\alpha)M(\alpha)\, dy(\alpha),$$

with $|\tau_i - \alpha| < \delta$ in each interval of integration. Now $X(\cdot)$ is uniformly continuous (in norm) and $\|X^{-1}(\cdot)\|$ is bounded on $[0, t]$. This implies that $\|X(\tau_i)X^{-1}(\alpha) - I\| \leqq \|X^{-1}(\alpha)\|\cdot\|X(\tau_i) - X(\alpha)\| \to 0$ for each i as $\delta \to 0$, so that the stochastic integral converges in norm to $\int_0^t M(\alpha)\, dy(\alpha)$. Thus, we take limits-in-norm on both sides of (5.11), and rearrange terms to give

$$\begin{aligned}\int_0^t H(t,\tau)M(\tau)\, dy(\tau) &= \int_0^t A(\tau)\left(\int_0^\tau H(\tau,\alpha)M(\alpha)\, dy(\alpha)\right) d\tau\\ &\quad + \int_0^t M(\tau)\, dy(\tau)\end{aligned} \tag{5.12}$$

with probability one. It follows from (5.12) that (5.5) provides a solution to (5.2).

The preceding arguments verify

THEOREM 5. *Let the integral equation* (5.2) *be given, and let the integrals appearing therein be understood as Bochner and stochastic integrals, respectively. Then* (5.5) *provides a solution to* (5.2); *for each* t, *this solution is unique up to an equivalence. If* H *is the Hilbert space generated by* $y(\cdot)$ *as described in Section* 2, $x(t) \varepsilon H$ *for each* t.

A direct computation, based on (5.3) and (5.5), shows the covariance matrix of $x(\cdot)$ to be

$$P(t, s) = \int_0^{\min(s,t)} H(t, \tau)M(\tau)M^*(\tau)H^*(s, \tau)\, d\tau. \tag{5.13}$$

Since $H(u, t)H(t, s) = H(u, s)$, (5.13) can conveniently be rewritten as $P(t, s) = H(t, s)P(s, s)$ whenever $t > s$. Thus, if $P(s, s)$ is invertible for all $s > 0$, $R(t, s) = H(t, s)$ for $t > s$. It follows from this and (5.6) that such a normalized covariance satisfies the wide-sense Markov criterion (3.4). The prediction $\hat{E}[x(t+\alpha) \mid x(s), s < t]$ is then given immediately as $X(t + \alpha)X^{-1}(t)x(t)$, which means that the minimum-mean-square-optimum linear predictor is composed solely of summers and multipliers.

It remains to investigate the case of singular $P(s, s)$. We note first that, if $M(\cdot)$ is nonsingular in some neighborhood of the origin, $P(s, s)$ is nonsingular for every positive argument. Also, if $P(s, s)$ is nonsingular and $t > s$, $P(t, t)$ is nonsingular. Both these facts are demonstrated below; they grow out of the proof that, whether $P(s, s)$ is nonsingular or not, $x(\cdot)$ is always MWM.

Consider the integral $W(t) = \int_0^t [X(\tau)]^{-1}M(\tau)M^*(\tau)[X^*(\tau)]^{-1}\, d\tau$. Clearly, $W(\cdot)$ is Hermitian, and may be written in the canonical form $C_t^* D_t C_t$, where C_t, D_t, and E_t are all defined analogously to the C, D, and E occurring after (2.14). Then $R(t, s)$, $t > s$ is (uniquely) specified by

$$R(t, s) = X(t)[C_s^* D_s E_s^{-1} C_s]X^{-1}(s). \tag{5.14}$$

To verify the assertion that $R(t, s)$ is given by (5.14), we show that $P(t, s) - R(t, s)P(s, s) = 0$ is satisfied. The calculation follows when we substitute $P(s, s) = X(s)W(s)X^*(s) = X(s)[C_s^* D_s C_s]X^*(s)$, and note that $C_s^* C_s = I$ and $D_s E_s^{-1} D_s = D_s$.

To check the MWM property, we introduce (5.14) into (3.4). The resulting matrix equality remains unchanged if both sides are pre- or post-multiplied by invertible matrices. Therefore the MWM property is equivalent to requiring that

$$P_s = P_t P_s \qquad t > s \tag{5.15}$$

where we define $P_s = C_s^* D_s E_s^{-1} C_s$. A simple check shows that P_s is idempotent and Hermitian, and therefore a projection. Then (5.15) is true if and only if $P_s \subset P_t$ (for the meaning of this relation and associated concepts, see [5], pp.

41–49). Thus $x(\cdot)$ is indeed MWM if we show that $\mathfrak{M}_s \subset \mathfrak{M}_t$ (for $s < t$), where the $\mathfrak{M}$'s are subspaces associated with the corresponding projections.[5]

If c_{si} is the ith column vector of C_s^*, $W(s)c_{si} = \lambda_{si}c_{si}$. At the same time, $P_s c_{si} = c_{si}$ or zero according as $\lambda_{si} \neq 0$ or $= 0$. Because the c_{si} are mutually orthogonal vectors, $i = 1, 2, \cdots, n$, $W(s)v \neq 0$ implies $P_s v \neq 0$ and conversely, for any vector v. It follows that $\mathfrak{M}_s$ is precisely the subspace spanned by those vectors v for which $W(s)v \neq 0$. Hence our proof of $\mathfrak{M}_s \subset \mathfrak{M}_t$, $s < t$, consists of showing that $W(t)v \neq 0$ whenever $W(s)v \neq 0$.

Consider first $v = c_{si}$, $\lambda_{si} \neq 0$. Then $(c_{si}, W(s)c_{si}) = \lambda_{si}$. But, using the definition of $W(\cdot)$, and taking $G(\tau) = [X(\tau)]^{-1}M(\tau)$, we have $(c_{si}, W(s)c_{si}) = \int_0^s \|G(\tau)c_{si}\|^2 \, d\tau$. This integral cannot then be zero, so that it is in fact positive. We have thus proved the λ_{si} to be nonnegative. More important, we have the inequality

$$(5.16) \qquad (c_{si}, W(t)c_{si}) = \int_0^t \|G(\tau)c_{si}\|^2 \, d\tau \geqq \int_0^s \|G(\tau)c_{si}\|^2 \, d\tau > 0$$

which shows that $W(t)c_{si}$ is not zero.

The above arguments are readily applied to a general vector v. We write $v = \sum a_i c_{si}$, and note that $W(s)v = 0$ unless, for at least one i, a_i and λ_{si} are both nonzero. Therefore, $W(s)v \neq 0$ implies $(v, W(s)v) = \sum |a_i|^2 \lambda_{si}$ to be strictly positive. As before, $(v, W(\cdot)v) = \int_0^{\cdot} \|G(\tau)v\|^2 d\tau$. We may then apply (5.16), with v replacing the c_{si} appearing there. This shows $W(t)v \neq 0$, and completes the proof that $x(\cdot)$ is MWM.

Since $X(\cdot)$ is invertible, $G(\cdot)$ is singular if and only if $M(\cdot)$ is singular. If, however, $M(\cdot)$ is nonsingular in some neighborhood of the origin, $\int_0^\delta \|G(\tau)v\|^2 d\tau > 0$ whenever v is nonzero and $\delta > 0$. It follows from what has been proved that $W(\delta)v \neq 0$ and, *a fortiori*, that $W(t)v \neq 0$ for any $t > 0$. Hence $W(t)$ is nonsingular for every $t \neq 0$, from which follows also that $P(t, t)$ is nonsingular.

We summarize the above proofs in

THEOREM 6. *Let $P(\cdot,\cdot)$ be specified by* (5.13). *Then $x(\cdot)$ is* MWM, *with an $R(\cdot,\cdot)$ given by* (5.14). *If we define $t_0 = \inf\{t: M(\cdot)$ is nonsingular on a set of positive measure in $[0, t)\}$, then $P(t, t)$ is nonsingular for all $t > t_0$ (but the converse fails to hold). Further, if $P(s, s)$ is nonsingular, and $s < t$, then $P(t, t)$ is nonsingular. Finally, if $P(s, s)$ is nonsingular,*

$$(5.17) \qquad R(t, s) = H(t, s), \qquad t > s.$$

The theory just presented may be useful in many applications. For instance, $M(\cdot)$ could be a switching function in a multivariable control system, in which it is desired to interpolate an estimated value of $x(t)$ between samples. If the input to such a system is not white noise, but rather some integral thereof (e.g.,

[5] The vector space to which we refer in this and the following paragraphs is the ordinary n-dimensional vector space, in which the inner product is the vector dot product, the norm of a vector its (Euclidian) length, etc.

a Wiener process), the prediction can still be performed by artificially enlarging the dimension of $A(\cdot)$. The details of this procedure are left to the reader.

It has been pointed out before that optimum estimates of functionals such as the $z(t)$ given in (4.11) can be found. This, of course, applies *pro tanto* to a $z(t)$ derived from a differential equation. Indeed, if $K(t, \tau)$ is of the form $F(t)e^{B\tau}$, if $M(\cdot)$ is a constant nonsingular matrix, and (5.2) a constant-coefficient equation the exponential of whose kernel commutes with B, the integrands appearing in (4.11) are exponential, and may be directly evaluated.

REFERENCES

[1] Bellman, R. (1953). *Stability Theory of Differential Equations.* McGraw-Hill, New York.

[2] Beutler, F. On multivariate prediction. To be published.

[3] Doob, J. (1944). The elementary Gaussian processes. *Ann. Math Statist.* **15** 229–282.

[4] Doob, J. (1953). *Stochastic Processes.* Wiley, New York.

[5] Halmos, P. (1950). *Introduction to Hilbert Space and the Theory of Spectral Multiplicity.* Chelsea, New York.

[6] Hille E. and Phillips, R. (1957). *Functional Analysis and Semi-Groups.* (rev. ed.). Amer. Math. Soc. Colloquium Publications **31.**

[7] Wang, M. W. and Uhlenbeck, G. E. (1945). On the theory of the Brownian motion II. *Rev. Mod. Phys.* **17** 323–342.